THE RELIABILITY OF THE IDENTIFICATION OF THE HUMAN VOICE

By

FRANCES McGEHEE

A Dissertation

Submitted to the Board of University Studies of the Johns Hopkins University in Conformity with the Requirements for the Degree of Doctor of Philosophy

1936

Reprinted from *The Journal of General Psychology*, 1937, 17, 249-271.

READERS IN VOICE EXPERIMENTS

Men	Men	Women
Alexander	Lydane	Bailey
Bright	Mahon	Bowers
Brintnall	Morphopoulos	Bridges
Bromiley	Phillips	Cowles
Bunyan	Rosenthal	Harris
Clark	Severn	Huebner
Dorcus	Squire	Jennings
Doyle	Tsiang	Kagey
Durand	Wilder	Massie
Durding	Wilson	Mast
Gardner	Wright	Medinger
Glantz	Young	Neuman
Harrison	Zeigler	Preece
Hirschfelt		Rau
Hoyt		Rucker
Kahn		Throckmorton
Kanne		Watkins
Leatherman		Williams

THE RELIABILITY OF THE IDENTIFICATION OF THE HUMAN VOICE*[1]

The Johns Hopkins University

FRANCES MCGEHEE[2]

SECTION I: INTRODUCTION

I. THE ORIGIN OF THE PROBLEM

The reliability of court procedure with reference to identification of voices is open to serious question from a psychological point of view. Investigations by psychologists on fidelity of report have not included a study of voice identification. Attention was forcibly directed to a study of this kind in January, 1935, at the time of the world-famous trial of State v. Hauptmann.

The court procedure followed in this case with regard to voice identification, although consistent with established legal precedent, was questioned from a psychological standpoint. Sworn testimony of positive identification of the defendant by his voice almost three years after the offense was committed, was accepted by both courts as material evidence against the defendant. The argument for the defense was that the "verdict was against the weight of evidence: Colonel Lindbergh's identification of the defendant by his voice was entitled to little weight" (13).

Assuming that it is possible to recognize an unfamiliar voice after

*Recommended for publication by Roy M. Dorcus, and received in the Editorial Office on October 16, 1936.

[1]A dessertation submitted to the Board of University Studies of The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy, June, 1936.

[2]The writer wishes to acknowledge her indebtedness to Dr. Roy M. Dorcus of The Johns Hopkins University under whose guidance the investigation was planned; to Dr. Leon Sachs of the Political Science Department of The Johns Hopkins University, and to Dr. John Strahorn, Jr., of the University of Maryland Law School, for suggestions concerning the case material presented; to those members of the staff of the St. Mary's Seminary, Randolph-Macon Women's College, Loyola College, Notre Dame College of Maryland, Maryland College for Women, University of Baltimore, and The Johns Hopkins University, who made it possible to secure the subjects for the experiments; and finally to the readers who gave generously of their time in carrying out the experiments.

R. UNIV. BIBLIOTHEEK LEIDEN

a considerable interval of time has elapsed since its first occurrence, to what extent or with what frequency are unfamiliar voices recognized? The answer to this question should assist us in determining how much weight may be attached to the recognition of a voice as evidence.

II. Legal Aspects: Its Bearing upon Testimony

A. Admissibility of voice identity as evidence.

From legal sources, a few cases will be cited concerning the admissibility of voice identity as evidence. It is generally conceded that a person may be recognized and identified by his voice, if the hearer is acquainted with the speaker's voice (6). But, if there is no recognition of voice it is improbable that sufficient evidence can be supplied to authenticate the antiphonal speaker. Usually the question of voice recognition arises in connection with telephone conversations or with offenses committed in the darkness, in which case the voice of an unseen person is almost the only clue to his identity.

A telephone conversation, unless otherwise objectionable, is admissible as evidence provided the identity of the person communicated with is satisfactorily established (5). The test of authentication universally accepted is the recognition of a familiar voice and in some cases, the subsequent recognition of a voice unfamiliar at the time of the conversation. Its force would depend on whether the jury believed the witness. Even though the witness is not identified positively as the person speaking, the telephone conversation may nevertheless be received in evidence, the uncertainty of recognition going only to the weight of evidence (9). It is of course recognized that the mechanism of the telephone is sufficiently reliable so that the mere fact that the conversation was so conducted should not make it inadmissible.

In a contract action, the trial court admitted testimony of the plaintiff's agent that he had made the agreement over the telephone with one claiming to be the defendant; that he had never heard the defendant's voice on any other occasion until trial two months later, when he recognized it as the one he had heard over the telephone. The opinion of the higher court was that the evidence should have been excluded (4).

In another case, the testimony of a witness, not knowing plaintiff,

nor his voice at the time of a telephone conversation, but upon hearing the voice at the trial a year later, identified it, was admitted, but considered far-fetched and of a type easily manufactured (19).

Where an offense is committed in the darkness, a witness who did not distinguish the features, but did hear the voice of the one who committed the offense, may testify that the defendant was the offender and that he recognized him by his voice (14). This mode of identification, however, has never been extended to telephone conversations (8). For such an extension, at least in the absence of some marked peculiarity in the speaker's voice, would seem undesirable in view of the uncertainty which as a matter of common experience attaches to the recognition of even a familiar voice when heard over the telephone (4).

Considerable complaint was made of the identifying testimony given by a mother and her daughter, upon whom assault was made, when they testified that they recognized form and clothes and also identified defendant, as the one in the room, by his voice. The opinion of the court was that a witness may testify that he recognized the accused by his voice, since that is the principal and most reliable identifying evidence of crimes committed in the darkness, or upon persons who are blind (14).

Witnesses in a case (18) tried in 1861 testified that they identified the defendant's dog as one of two dogs that killed sheep. They said the dog had a "very coarse voice," that they could identify it by its bark. The question for the jury to determine was whether the dog had such a peculiar bark or voice that witnesses acquainted with the animal could satisfactorily identify it, by only hearing it bark in the night without seeing it. The decision was that some persons have such peculiar voices that they can be identified by acquaintances, who hear them talk, without seeing them. It seemed reasonable, therefore, that some dogs may bark in such a manner and have such singular voices that they can be identified in the night by persons who know them well, by merely hearing them bark, without seeing them.

From the legal standpoint, therefore, a person may be identified by his voice as well as by his lefthandedness, his tatoo marks, his skill in chemistry, his residence in Jonesburg, his personal history at college or in a shop.

B. Development of Scientific Methods in Evidencing Identity.

Wigmore (17) has summarized the significance of certain modes of identification, in the statement that the use of scientific methods by experts has developed data in the logic of evidencing identity which are vastly in advance of the simple data formerly procurable by the unaided senses. Assuming the witnesses to be reliable, the data often amount to proof, not merely to admissibility.

III. Psychological Aspects: Fallibility of Memory

A. Errors in sworn testimony.

One of the obvious weaknesses of testimony is due to the fallibility of memory. Laboratory experiments have been devised for measuring the degree of fidelity of report which can be expected under good circumstances. Legal testimony, however, seldom rests upon experiences so favorable for correct remembrance as the events arranged in an experiment. The immediate recall precludes all the weakening and distortion of the memory that occurs with the passage of time.

An experiment by Boring (2) called for a report on a motion picture scene lasting one minute. The observers were not told that they would be subjected to a memory test, but were instructed to pay careful attention to the picture. In addition to reporting the items remembered, each witness indicated those of which he was certain, and those he would be willing to swear were correct. Under circumstances such that the maximum of attention is guaranteed and with the report called for immediately after the event, the best group reported only 42.1 per cent of the facts and of these only 32.6 per cent were correct. Under oath the per cent of items reported fell to 26 per cent, but of these 22.6 per cent were right.

Whipple (15) concluded from data which he assembled that, although errors are reduced about half by the introduction of an oath, there still remains about 10 per cent error in sworn testimony.

Breukink (3) was interested in the factors affecting fidelity of report, especially practice. He found that if the individual's report of different pictures presented on occasions a week apart, were divided into sections, the earlier parts were more reliable than the later parts. Thus the items first coming to mind were more likely to be correctly reported. Reliability also increased with practice, the

third picture being more reliably reported than the second, and the second than the first. Practice increased the reliability of oath likewise. No consistent sex differences were found in fidelity of report.

B. Errors in Identification of Sounds.

In the case of hearing it is frequently extremely difficult to identify sounds as of a certain character and often impossible to assign them to their causes. The direction from which sounds come is subject to as much misinterpretation as their quality and intensity. Münsterberg (11) in his regular psychology course at Harvard had several young men, mostly between twenty and twenty-three years of age, take part in an experiment in which they were asked to describe the sound that they would hear and to say from what source it came. The sound which he produced was the tone of a large tuning-fork, which he struck with a little hammer below the desk, invisible to the students. Exactly two students of the first hundred whose papers were taken up at random, recognized it as a tuning-fork tone. All the other judgments took it for a bell, or an organ-pipe, or a muffled gong, or a brazen instrument, or a horn, or a 'cello string, or a violin. Or they compared it with as different noises as the growl of a lion, a steam whistle, a fog-horn, a fly-wheel, a human song. The description, on the other hand, called it: soft, mellow, humming, deep, dull, solemn, resonant, penetrating, full, rumbling, clear, low; but then again, rough, sharp, whistling and so on. Considering the conditions under which these results were obtained as being rather uniform in that everyone knew beforehand that he was to observe the sound stimulus, which was announced by a signal, how much more would the judgments have differed if the tone had come in unexpectedly?

C. Intentional Memory and Incidental Memory.

Memories involved in testimony are almost always of the so-called incidental or unintentional sort. In many cases there is no definite intention to observe events and to remember them. Both Meyers (10) and Shellew (12) found from studies of individual differences in incidental memory, that it is extremely weak and unreliable in comparison with intentional memory. The differences between incidental memory and intentional memory for various kinds of material were measured by Achilles (1) under experimental

conditions that were far more favorable to good incidental memory than entirely uncontrolled experiences, and free from emotional disturbances and shock. The incidental memory for syllables, expressed in terms of intentional memory was 81.3 per cent; the association of names with photographs was 57.2 per cent. It is suggested, in the light of such studies, that all values for intentional memory in testimony should be reduced at least one half to indicate their incidental memory value. Tests made on naive and unsuspecting observers render this estimate conservative.

D. Influence of Time-interval upon Memory.

The figures in Table 1 from Dallenbach (6) seem to be fairly typical of the dependence on time-interval when all conditions are

TABLE 1 (*from Dallenbach*)
INFLUENCE OF TIME INTERVAL UPON MEMORY FOR NARRATIVE MATERIAL

Interval	Items recalled	Per cent error
Immediate	765	10.5
5 days	735	14.3
15 days	750	18.0
45 days	569	22.4

favorable for correct memory. The accuracy decreases much more rapidly than the quantity of material reported. The error increases more than 100 per cent in the course of 45 days. Such an interval is not a long one in legal affairs. The additional distortion that would be introduced through the frequent repetition of testimony before a coroner's jury, the grand jury, the trial jury, through the conferences with lawyers and through the emotional excitement attending all these circumstances, can only be surmised.

IV. SUMMARY

The foregoing survey merely gives the setting for the problem of voice recognition and is by no means a history, for up to the present time, no work has been reported in the psychological literature which bears directly upon the recognition of voices. It is, therefore, a new problem upon which no experimentation has been made. Consequently it was necessary to devise a technique and numerous procedures especially adapted to the nature of the problem as the investigation progressed. In the discussion of the results, reference will be made

to the works cited above which have been given to indicate the angles of approach and to define the limits within which the investigation is centered.

SECTION II: EXPERIMENTAL INVESTIGATION OF VOICE RECOGNITION

I. Questions Arising within the Problem

Thirty-three situations were devised in an endeavor to answer the following questions concerning the recognition of voices:

1. To what extent are unfamiliar voices recognized after varying intervals of time?
2. What is the effect of increasing the number of voices heard in a series, in which occurs an unfamiliar voice to be subsequently recognized?
3. To what extent is an unfamiliar voice recognized as having been heard previously when disguised by a change in pitch?
4. To what extent are unfamiliar voices of different nationalities subsequently recognized?
5. Are there sex differences in recognition of voices?
6. Are men's voices or women's voices recognized more frequently?
7. What is the influence of the time-interval upon memory for voices?

II. Procedure

The general procedure was that of a group experiment. Each member of a group was given a sheet of instructions which were as follows:

> You will hear some sentences read behind the screen. Listen carefully as you will be tested later. Do not write anything but your name and date in the space indicated.

A paragraph of 56 words was read once behind a screen placed inside the room so that the reader could enter and leave unobserved by the auditors. The reader was unknown to the group. No particular qualifications were considered as important for the reader, except the absence of a peculiar dialect or a noticeable speech defect.

III. The Subjects

There were two kinds of subjects: readers and auditors. Usually the readers were graduate students. In every case in which the auditor knew the reader, that subject's judgment was discarded. Thirty-one male voices, including five foreign voices, and 18 female voices were employed in the experiments.

The auditors were groups of students at two preparatory schools, a Catholic seminary, two colleges for women; undergraduates at two universities and a group of law students. The number and character of the groups subjected to the voice recognition tests are given in Table 2. In some of the institutions, several groups were

TABLE 2
Auditors

Number	Men
270	Johns Hopkins University
34	University of Baltimore
135	St. Mary's Seminary
68	University of Maryland Law School
47	Loyola
554	total number of men
Number	**Women**
99	Maryland College for Women
39	Randolph-Macon Woman's College
48	Notre Dame of Maryland
186	total number of women
740	total number of auditors

tested, so that the auditors comprised 15 groups. In eight of these, two or more experiments were made with the result that altogether, 1404 judgments were obtained, related to various parts of the general problem.

All of the groups were tested for voice recognition after one or after two of the following time intervals:

1 day	1 week	1 month
2 days	2 weeks	3 months
3 days	3 weeks	5 months

IV. Number Selection

A second set of instructions was given to each member of the group on the day of the tests:

> You will hear five voices behind the screen, one of which you have heard before. Indicate your selection of the repeated voice by putting a check mark after one of the following numbers which correspond to the order in which the voices are heard.
> 1 ..... 2 ..... 3 ..... 4 ..... 5 .....

The selection used previously was read successively by five individuals unknown to the group and concealed from them by a screen. If, for example, the fourth reader was the one whose voice was heard on the first day of the experiment, an interval of a week having elapsed, every listener who recognized the fourth voice as the one he heard previously would check number *4* on his instructions sheet. The order in which the original voice was heard on the second day of the experiment varied from group to group.

The third or middle position was never assigned to the reader whose voice was to be recognized, since it was assumed in the beginning that when the numbers *one* to *five* inclusive, were presented in a serial order, the middle number (3) would be selected most frequently by a large group of individuals, provided there was no basis for their selection other than chance. The assumption was later verified by an item added to the instruction sheets distributed to four groups on the first day of their experiment.

> Put a check mark after one of the following numbers:
> 1..... 2..... 3..... 4..... 5.....

The results of the number selection made by 189 men and 155 women are shown in Table 3. The results for the men closely ap-

TABLE 3
Number Selection

Numbers in series	Percentage of 189 men selecting number	P. E.	Percentage of 155 women selecting number	P. E.
1	8%	1.9	16%	2.9
2	17%	2.7	16%	2.9
3	44%	3.6	40%	3.9
4	20%	2.8	17%	3.0
5	11%	2.2	11%	2.5

proximate the form that has been called the normal probability curve. The number distribution for the women is less symmetrical, although, in both groups, three is the preferred position. On the basis of chance alone, one would expect each number to be selected by 20 per cent of the group. The spatial presentation evidently influences the selection. An extra-chance factor seems to be operative, which we may designate as a number habit.

V. Variations from the General Procedure

A. Increase in the number of voices heard initially.

An element of confusion was introduced in the procedure employed with five groups by increasing the number of voices heard on the first day of the experiment. The time interval, two days, was constant for each group.

B. Missing voice.

A modification of the general procedure was used on four groups, in order to find out whether the judgments of the groups were made in accordance with the instructions or whether they simply checked a number at random. The judgment, when the general procedure was employed, consisted in checking one of five numbers in a series to represent the order of occurrence of a voice which had been heard on a previous day. According to the modified procedure, the original voice was to be designated by its *absence* from a second series of voices which immediately followed the first series. A third set of instructions was given to four groups (*I, J, K, L*) immediately after the usual test of voice recognition:

> You will hear only four voices behind the screen, in the same order in which they have been heard, except that one of the five voices will be missing. Encircle a number in the preceding series from 1 to 5 to represent the voice which *does not occur* this time.

The voice missing from the second series was the one heard originally. Everyone who actually recognized the recurring voice in the first series, should have noticed that the familiar voice was missing from the second series of four voices. The results of four groups selecting the order of a recurring voice and selecting the order of a missing voice are compared in Table 4.

The percentage for each group is less in the case of the missing

TABLE 4

PERCENTAGE OF AUDITORS IN FOUR GROUPS SELECTING THE ORDER OF A RECURRING VOICE AND SELECTING THE ORDER OF A MISSING VOICE

Groups	Per cent of auditors selecting order of *recurring* voice	P. E.	Per cent of auditors selecting order of *missing* voice	P. E.
I	57%	12.0	53%	12.8
J	51%	1.7	48%	5.4
K	99%	1.1	90%	3.6
L	98%	1.3	67%	4.7

voice than in the selection of the recurring voice. The largest decrement occurred in Group *L* (98 women). A partial explanation which may be offered is the fact that there was a spontaneous acclamation by the majority of the students when they heard the repeating voice (*B*) which happened to be the fifth and last one in the series. It is suspected that those students who were doubtful, or had not formed their judgment quickly, made use of the cue afforded by the applause. This suspicion is strengthened by the fact that only two women selected a voice other than the fifth one.

The same sort of thing occurred in one other group (*K*, 69 men studying for the priesthood). The same voice (*B*) was the last in the series and the one to be recognized. There was a break in discipline such as occurred in Group *L*. Of the 68 *men in Group K*, 67 checked the last number in the series. In order to eliminate the cue difficulty and to provide a check on the results, the series was repeated with the original voice missing. In this case ten per cent failed to select the order of the missing voice.

C. Memory of the selection heard.

The modified procedure required the reading of the selection once on the initial day of the experiment and nine times on the second day. It should be remembered that the instructions given to each group on the first day were to listen carefully to the sentences which would be read behind a screen, for a test would be given later. The implication was that everyone should remember verbatim or at least the meaning of the selection heard. In this way the auditory attention of the group to the verbal selection was insured, without disclosing the nature of the experiment which was to follow. In many cases in which a person is called upon to identify an individual by his voice,

there is no warning or fore-knowledge that at some future time he will be required to identify a voice as one heard on a previous occasion.

On the second day of the experiment, the auditors were to select the voice which they had heard on the first day of the experiment. It was only natural that the attention of the group was centered now upon the *voices,* rather than upon the *words* spoken. After judgments had been made of the occurrence of a voice in one series and its absence in another series, the groups were instructed verbally to write from memory the selection which they had heard nine times that day and once intentionally on a previous day. The instructors in two of the groups misinterpreted the instructions given and insisted that the students write nine times the selection which they had just heard! No one was able to reproduce correctly the paragraph of 56 words. Everyone who attempted to write something got the idea that it had to do with moving pictures. The selection follows (7):

> The enormous success of moving pictures is based upon their power to arouse an illusory perceiving of action. The screen, although actually in darkness many times per second, is able to stimulate the audience much as do the continuous movements of actors in person—arousing and holding attentive attitudes, touching off emotional tendencies, exciting thinking behavior.

D. Disguised voice.

A further modification of the general procedure was that of having a reader disguise his voice by changing its pitch. He later resumed his natural voice, at which time the auditors were required to select one of a series of natural voices, as being the voice previously disguised.

E. Methods of making judgments.

The groups were asked at the conclusion of the experiment whether they made their judgment of the recurring voice as soon as they recognized it (method *B*), or after they heard *all* of the voices (method *A*).

F. Attested recognition of a voice.

In one experiment, a group was asked whether they would be willing to swear that they recognized one of the voices which they had heard on only one other occasion. Fourteen out of a group of

50 undergraduate men, i.e., 28 per cent, were willing to swear that they recognized the voice. One of them was wrong.

G. Sex differences in recognition of voices.

Sex differences were studied under four conditions:

I	II	III	IV
Male auditors	Male auditors	Female auditors	Female auditors
Male voices	Female voices	Female voices	Male voices

H. Foreign voices.

Under Condition I, five voices representing four nationalities were heard. In this case the voices were both unfamiliar and strange, any one of which was easily distinguishable in a group of American voices. The voice to be recognized was that of a German student who had been absent from Germany six months. His voice was the second heard in a series, composed of another German, who had been in this country the same length of time; a Chinese, in residence 5 years; a Greek, 9 years; and a Russian, 13 years. It was surprising that the voice most frequently confused with the German voice, was not the other German voice, but the Chinese. The following nationalities were attributed to the original voice, besides the correct "German": Japanese, Norwegian, Dutch, Jewish, Assyrian, Polish-Russian, Swedish, Russian with French influence, Spanish and German-Jew.

I. Criticisms of the experiment.

Criticisms of the experiment were sought from the group of 68 law students after they had been tested for recognition of a foreign voice. Two statements represent the extremes of opinion:

> "The test was not fair because the voices were too dissimilar, so that it was easy to pick out the repeating voice."
>
> "It would not surprise me if they were all made by the same person."

The latter statement was to be expected since there is usually at least one wary individual who suspects trickery in a psychological experiment.

VI. Interpretations of Results

A. Influence of time-interval upon voice recognition.

Table 5 shows the percentage of auditors recognizing an unfamiliar voice after increasing intervals of time. As the time-interval

TABLE 5

VOICE RECOGNITION AFTER INCREASING INTERVALS OF TIME

Male auditors and male voices.

Time interval Days	Percentage of auditors recognizing a voice	P.E.	Groups
1	83.0%	4.6	*B, G*
2	83.0%	10.8	*F*
3	81.0%	4.7	*H*
Weeks			
1	80.8%	2.9	*D, E, L*
2	68.5%	3.9	*C, K*
3	51.0%	5.4	*J*
Months			
1	57.0%	12.0	*I*
3	35.0%	6.0	*K*
5	13.0%	5.2	*M*

increases from one day to three months, there is a general trend towards a decrease in percentage of auditors who are able to recognize an unfamiliar voice the second time it is heard. The tests of voice recognition after intervals of time, like the results obtained by Dallenbach (6) on memory for narrative material, indicate the influence of time-interval upon memory.

B. Effect of increasing the number of voices heard initially.

Table 6 shows the effect of increasing the number of voices heard initially, on the percentage of auditors who recognized one of the voices subsequently. When *one* voice was heard on the first day and five voices two days later, 83 per cent of the auditors in Group *F*, selected the second voice of the series as the one heard previously.

TABLE 6

VOICE RECOGNITION AS AFFECTED BY INCREASING THE NUMBER OF VOICES IN THE INITIAL SERIES

Group	Number of voices in initial series	Time intervals 2 days	P.E.
F	1	83%	10.8
C	2	77%	5.0
M	3	94%	3.5
K	4	52%	6.2
N	5	46%	7.1
		1 week	
A	5	8%	7.5

When *two* voices were heard on the first day, one of which occurred in a series of five, two days later, 77 per cent of the auditors in Group *C* selected the recurring voice. When *five* voices were heard on the first day, only one of which occurred in a series of five voices two days later, 46 per cent of the auditors in Group *N* selected the recurring voice. Thus for a given time-interval, the effect of increasing the number of voices in the initial series, was that the percentage of auditors who subsequently recognized one of the voices when it recurred in a new series of voices, was decreased.

When five voices were presented at the beginning of an experiment and one of them was included in another series of five voices presented after an interval of a week, only eight per cent of the auditors in Group *A* recognized the voice which occurred in both series.

C. Effect of disguise upon voice recognition.

Table 7 shows the percentage of auditors in two groups, *K* and *F*, which recognized an unfamiliar voice as having been heard previously

TABLE 7

RECOGNITION OF A VOICE PREVIOUSLY DISGUISED BY A CHANGE IN PITCH

Male auditors and male voices.

Group	Time interval	Percentage of group recognizing disguised voice subsequently
K	Same day	61% 5.4 P.E.
F	One day	67% 3.9 P.E.
		Same voice undisguised
B	One day	80% 5.6 P.E.

disguised by a change in pitch. The same reader disguised his voice for both groups. In another group, *B*, 80 per cent of the auditors recognized the voice after one day when it was heard on both occasions in its natural pitch. A comparison of this percentage with other percentages presented in Table 7, indicates that a voice disguised by a change in pitch was recognized the following day when it was heard in its natural pitch, by a smaller percentage of auditors than when it occurred on both occasions in its natural pitch. The difference in recognition of a voice in its natural pitch and in an unnatural pitch amounted to 13 per cent.

D. Recognition of a voice of foreign nationality.

Table 8 shows the percentage of auditors (68 Law students) which selected one of a series of five foreign voices as the voice which they

TABLE 8

RECOGNITION OF A FOREIGN VOICE OCCURRING IN A SERIES OF FOUR OTHER FOREIGN VOICES

Time interval: two days. Male auditors and male voices.

Order and nationality of voices heard by Group *O*	Number of auditors selecting voice heard two days previously	Percentage of total group
1. German	2	3%
2. German (heard previously)	55	81%
3. Chinese	8	12%
4. Greek	1	1%
5. Russian	2	3%

had heard two days previously. The procedure used, the time-interval, the order of the recurring voice in a series and the conditions under which those in Group *O* were tested, were identical with those of Group *F*, except that Group *O* heard foreign voices, instead of ordinary voices. The difference in results when the two groups are compared should indicate, therefore, the difference in recognition of a foreign (German) voice and an ordinary American voice.

In Group *F*, 83 per cent of the auditors selected the second voice in a series of five unfamiliar voices, as the one which they had heard two days previously. In Group *O*, 81 per cent selected the second voice in a series of five foreign voices as the one which they had heard two days previously. Since, under the conditions employed in the two experiments, the frequency of recognition of both types of voices is approximately the same, it may be inferred that results obtained on recognition of foreign voices after longer periods of time, would bear out the results already obtained on unfamiliar American voices.

E. Sex differences in voice recognition.

Table 9 shows the percentage of male auditors and the percentage of female auditors who recognized voices under comparable conditions. These results rearranged in Table 10 indicate the difference in ability of men and women in the case of voice recognition. Under

TABLE 9

SEX DIFFERENCES IN VOICE RECOGNITION

Auditors	Interval	Percentage of auditors recognizing *male* voices	P.E.	Percentage of auditors recognizing *female* voices	P.E.
Men	2 days	83%	10.8	94%	3.4
Women	2 days	46%	7.1	46%	7.9
Men	1 week	62%	6.8	97%	2.8
Women	1 week	98%	1.4	47%	7.9

TABLE 10

EXTENT TO WHICH MEN AND WOMEN DIFFER IN ABILITY TO RECOGNIZE VOICES

Group	Percentage of male auditors recognizing voice	P.E.	Group	Percentage of female auditors recognizing voice	P.E.
F	83%	10.8	*N*	46%	7.1
L	94%	3.4	*D*	46%	7.9
B	62%	6.8	*L*	98%	1.4
E	97%	2.8	*D*	47.5%	7.9
Average	84%		Average	59.3%	

similar conditions for both sexes, a greater percentage of men recognized voices than did women. Men, however, recognized women's voices more frequently than they recognized men's voices. The percentage of voice recognition which was obtained under four separate conditions is shown in Table 11.

Are men's voices or women's voices recognized more frequently? Comparative percentages of recognition of voices representing both

TABLE 11

VOICE RECOGNITION UNDER FOUR CONDITIONS

Conditions	Percentage of group recognizing recurring voice
Male auditors; female voices	95.5%
Male auditors; male voices	72.5%
Female auditors; male voices	72.0%
Female auditors; female voices	46.7%

TABLE 12
RECOGNITION OF MALE AND FEMALE VOICES

Group	Percentage of auditors recognizing a recurring *male* voice	P.E.	Group	Percentage of auditors recognizing a recurring *female* voice	P.E.
F	83%	10.8	*L*	94%	3.4
N	46%	7.1	*D*	46%	7.9
B	62%	4.9	*E*	97%	2.8
L	98%	1.4	*D*	47.5%	7.9
Average	72.2%		Average	71.1%	

sexes are presented in Table 12. Since the difference between the averages of the two groups (72.2% for male voices; 71.1% for female voices) may be considered negligible, we may conclude that, under the conditions outlined in the general procedure, men's voices and women's voices, on the whole, are recognized with approximately equal frequency.

F. Reliability of the general procedure employed.

Table 13 shows the consistency of results obtained from male auditors hearing male voices under the conditions outlined in the general procedure after intervals of one day and two days. Group *B* and Group *G* were tested for memory of voices *L* and *Z*, respectively, after an interval of one day. There was a small variation in percentage of recognition from group to group and from voice to voice. The variation of 6 per cent may be attributed either to the difference in the two groups or to the difference existing between the two voices.

The same voice, *Z*, was recognized by 86 per cent of Group *G*

TABLE 13
RECOGNITION OF VOICES HEARD UNDER SIMILAR CONDITIONS BUT BY DIFFERENT GROUPS OF AUDITORS AFTER INTERVALS OF ONE DAY AND TWO DAYS
Male auditors and male voices.

Group	Voice	Percentage of group recognizing voice	P.E.
		Interval: one day	
B	L	80%	5.6
G	Z	86%	9.2
		Interval: two days	
O	R	81%	4.7
F	Z	83%	10.8

after one day and by 83 per cent of Group *F* after two days. The difference of 3 per cent in this case may be attributed to the increase in time-interval.

Since the results of both pairs of comparable groups are highly consistent for the two different time-intervals employed, the general procedure described earlier would seem to be reliable for the study of voice recognition.

C. Methods reported by auditors in the selection of the recurring voice.

Table 14 shows the methods employed by two groups of auditors in the selection of a recurring voice. Group *O* heard a series of five

TABLE 14
METHODS USED IN SELECTION OF RECURRING VOICE

Group	Method of judgment	Number of auditors using method	Percentage of total
O	A	45	67%
O	B	22	33%
		Total.....67	
K	A	51	80%
K	B	13	20%
		Total.....64	

	Of those indicating method *A*			
	Right		Wrong	
Group	No.	Per cent	No.	Per cent
O	35	78%	10	22%
K	29	45%	22	34%
	Of those indicating method *B*			
O	19	86%	3	14%
K	4	6%	9	14%

foreign voices. Group *K* heard a series of American voices, one of which had been disguised previously by a change in pitch. Two thirds of Group *O* heard all five voices before making a selection (method A). The remaining third made their selection as soon as they recognized the voice (method *B*). Of the latter, only three were wrong in their judgment, whereas 10 were wrong when the selection was made after hearing all of the voices.

Four-fifths of Group *K* heard all five voices before making a selection of the recurring voice. In 29 cases the judgment was correct; in

22 cases it was wrong. Of the remaining fifth who made their selection as soon as they thought they recognized a voice, nine were wrong, whereas 22 were wrong when all of the voices were heard before making a selection.

Method *A*, of hearing all of the voices before selection, was reported by approximately twice as many auditors as Method *B*. A smaller percentage of those reporting judgment by the latter method, however, were wrong than were those using the former method. The results presented on the two methods of making selections, reported by two groups, do not show conclusively which one of the methods is more reliable.

VII. Summary and Conclusions

The study of voice recognition, as has been stated, is a problem upon which no experimentation has been reported. Any conclusions, therefore, which the present data would seem to indicate should obviously be drawn with caution. In each case the conclusions offered are based upon a consideration of the experimental technique employed as well as the special procedures adapted to specific parts of the general problem. The following tentative conclusions are drawn concerning the recognition of unfamiliar voices:

1. The percentage of auditors recognizing an unfamiliar voice when it recurs in a series of other unfamiliar voices, decreases as the length of time increases between the first and second occurrence.

An unfamiliar voice was recognized by 83 per cent of a group of auditors when it occurred in a series of five unfamiliar voices on the day after the group first heard it. Another unfamiliar voice was recognized by 83 per cent of a group of auditors after an interval of two days. Not until the interval had been extended to two weeks, was there a notable decrement in the recognition of an unfamiliar voice. After three months, an unfamiliar voice was recognized by 35 per cent of those who heard it initially and by only 13 per cent, after five months.

2. The percentage of auditors, recognizing an unfamiliar voice after a given time interval, decreases, as the number of voices in a series in which the repeating voice first occurs, is increased.

When one of two voices heard by a group of auditors recurred two days later in a series of five unfamiliar voices, 77 per cent of the auditors recognized it. Only 46 per cent were able to recognize a

voice after two days when the number of voices in the initial series was increased to five. When the interval between occurrence and subsequent recognition was extended to one week, 8 per cent selected the recurring voice.

3. Disguising a voice by changing its pitch, also produces a decrease in the percentage of auditors who are able to recognize the voice when it recurs the following day in its natural pitch.

The difference in recognition of a voice in its natural pitch and in an unnatural pitch (disguise) amounted to 13 per cent.

4. An unfamiliar foreign (German) voice was recognized by approximately the same percentage of auditors as an unfamiliar American voice when each occurred under similar conditions. The German voice was the second in a series of five foreign voices, including another German, a Chinese, a Greek and a Russian. It was confused more frequently with the Chinese voice than with the other German voice.

5. Groups of male and female auditors, when contrasted in the extent to which they recognized voices under similar conditions, showed consistent differences.

(*a*). Men recognized women's voices more frequently than they recognized men's voices.

(*b*). Men recognized women's voices to a greater extent than women recognized men's voices.

(*c*). Men recognized women's voices with a frequency twice that of women recognizing women's voices.

(*d*). Men recognized voices as frequently as women recognized men's voices.

(*e*). Men recognized men's voices to a greater extent than women recognized women's voices.

(*f*). Women recognized men's voices more frequently than they recognized women's voices.

In general, men surpassed women in voice recognition.

6. Male and female voices which were heard under similar conditions by four different groups of auditors, were recognized with approximately equal frequency. The average percentage of the groups recognizing female voices (71.1%) and the average percentage of the groups recognizing male voices (72.2%) was taken as the basis for the comparison.

7. Results obtained from the present experiments on memory for voices, are in conformity with those reported by investigators of memory in general. The strength of the original voice as well as the strength of the material presented, decreases with time.

The reliability of court procedure in accepting testimony of positive identification of a defendant by his voice, in consideration of the length of time interval and the common fallibility of memory, would seem to be relatively low in the light of the present experiments designed to test the validity of such a procedure. The percentage of individuals who were *unable* to recognize an unfamiliar voice, increased from 17 per cent after one day to 87 per cent after five months. A great many factors enter into the identification of a voice when it is required in legal procedure, which yet remain to be attacked by experimental investigation. These factors require special consideration: (*a*) the conditions under which the voice of an unseen person is originally heard, i.e., whether over the telephone or, under more startling or unusual conditions, in the dark; (*b*) the strength of motivating factors which contribute to the degree of attention manifested by the auditors; (*c*) the manner in which the voice is heard, i.e., whether intentionally or incidentally, with or without reference to subsequent identification.

The present study has been confined to the voice recognition of normal adults. The plan is to extend the investigation to include other groups, i.e., the blind, the psychopathic, the mentally deficient and young children. Such an extension will require the development of a suitable technique adapted to the peculiarities and to the age level of the groups tested, and will also require the use of a method for precise voice analysis.

References

1. Achilles, E. M. Experimental studies in recall and recognition. *Arch. Psychol.,* 1920, **44**, 1-80.
2. Boring, E. G. Capacity to report upon moving pictures as conditioned by sex and age. *J. Amer. Inst. Crim. Law & Criminol.,* 1916, **6**, 820-834.
3. Breukink, H. Ueber die Erziehbarkeit der Aussage. *Zsch. f. Ange. Psychol.,* 1910, **3**, 32-87.
4. *Columbia Law Rev.,* 1911, **11**, 182.
5. *Columbia Law Rev.,* 1925, **25**, 972.
6. Dallenbach, K. M. The relation of memory error to time interval. *Psychol. Rev.,* 1913, **20**, 323-372.

7. DASHIELL, J. F. Fundamentals of objective psychology. Boston: Houghton Mifflin, 1928. Pp. 404.
8. DEERING & CO. v. SHUMPIK. *Minn. Rep.,* 1897, **67**, 348.
9. HECKMAN v. DAVIS. *Okla. Rep.,* 1916, **56**, 483.
10. MEYERS, G. C. A study of incidental memory. *Arch. of Psychol.,* 1913, **26**, 1-108.
11. MUNSTERBERG, H. On the witness stand. New York: Doubleday Page, 1913. Pp. 24-25.
12. SHELLEW, S. M. Individual differences in incidental memory. *Arch. of Psychol.,* 1923, No. 64, 1-77.
13. STATE v. HAUPTMANN. *Atlantic Rep.,* 1935, **180**, 809-829.
14. STATE v. HERBERT. *Kansas Rep.,* 1901, **63**, 516.
15. WHIPPLE, G. M. Manual of mental and physical tests. Baltimore: Warwick & York, 1915. Pp. 383-408.
16. WIGMORE, J. H. A treatise on the system of evidence in trials at Common law: Vol. 4. Boston: Little, Brown. Pp. 585.
17. ———. Textbook on evidence. Chicago: The Foundation Press, 1935. Pp. 111.
18. WILBUR v. HUBBARD. *N. Y. Barb.,* 1861, **35**, 303.
19. WOODRUFF v. BENESCH. *N. Y. Suppl.,* 1920, **20**, 880.

Friends University
Wichita, Kansas

VITA

Frances McGehee was born in Lynchburg, Virginia, February 21, 1912. She received her elementary and high school education in Lynchburg. From 1929-33 she attended Randolph-Macon Woman's College and received a Bachelor of Arts degree with a major in psychology. During the years 1933-36, she was enrolled in the school of Higher Studies of the Johns Hopkins University as a candidate for the Doctor of Philosophy degree in psychology. She is a member of Sigma Xi and of the Southern Society of Philosophy and Psychology.